DISE

**N1 FY10 Research Project**
**Research Sponsor: Navy Recruiting Command**
**POC: Mr. John Noble, Head of Research**

# Matching Navy Recruiting Needs with Social Network Profiles Using Lexical Link Analysis


**Primary Researcher**

**Dr. Ying Zhao**

**408.252.7746**

**408.218.8484**

***Associate Researchers:***
***Dr. Douglas J. MacKinnon***
***Dr. Shelley P. Gallup***
***Research Associate Professors***

***Distributed Information Systems Experimentation (DISE)*** ***Naval Postgraduate School***

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **JAN 2010** | | **00-00-2010 to 00-00-2010** |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Matching Navy Recruiting Needs with Social Network Profiles Using Lexical Link Analysis** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Distributed Information Systems Experimentation (DISE),Monterey,CA,93943** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**Personnel and National Security: A Quantitative Approach (Unclass), 25-28 January 2010, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland**

14. ABSTRACT

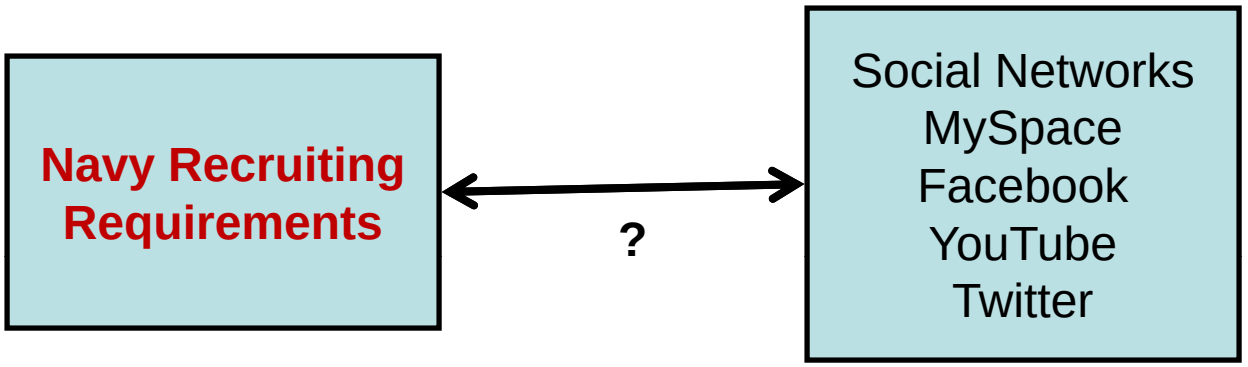
15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT **unclassified** | b. ABSTRACT **unclassified** | c. THIS PAGE **unclassified** | **Same as Report (SAR)** | **25** | |

DISE

# Research Issues/Scope

## Matching Navy Requirements in a Social Networking Environment




## Research Questions

- Do current service entry profiles match well with candidate profiles in the social networking environments?
- What is the methodologies facilitate matching profiles that could potential inform Navy Recruiting Marketing decisions?
- Are there sufficient matches to make marketing in this environment cost effective?

DISE

# Proposed Method

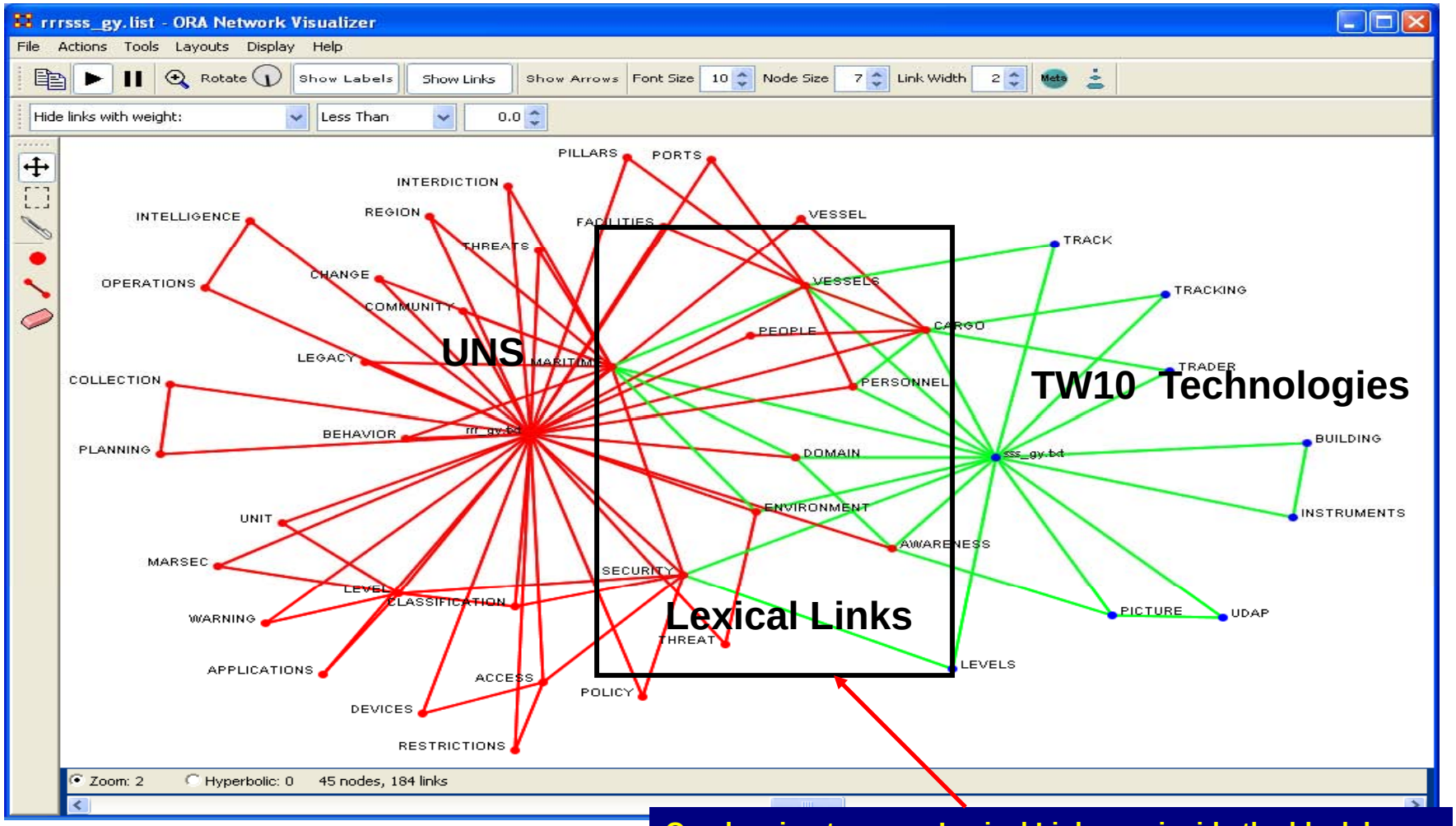
- **Lexical Link Analysis (LLA)**
  - Features – terms/keywords to describe Navy recruiting needs and candidate profiles
  - Link the features in context of dynamic social network environments, learn from on-going market places
- **Previous applications of (LLA)**
  - System self-awareness/ program-awareness: DISE demonstrated methodology within the Navy acquisition community
    - Matched Navy needs (e.g. need statements) with technologies (e.g. Trident Warrior technologies)
    - Resulting in improved acquisition decisions
    - (See next slide for example)

# Lexical Link Maps

## (From a previous application)

**Specifically: Matched Navy Needs Profiles to Trident Warrior (TW10) Technologies**

DISE




**Overlapping terms or Lexical Links are inside the black box**

DISE

# Navy Recruiting Research Tasks

**Select and Pre-Process data sets**

- Obtain Facebook data samples - 1,700 Facebook profiles site
- Obtain service entry requirement descriptions and extract features for service entry requirements
- Pre-process and extract features from a sample data set

**Apply Proposed Approaches**

- Profile the current service entry requirements using Lexical Analysis
- Profile Facebook sample data using Lexical Analysis
- Link profiles with rating descriptions using LLA generating potential recruiting candidates
- Develop mapping interface output matching qualified candidates

**Prepare Deliverables**

- NPS Technical Report submitted at the completion of the project
- Presentation at Annual Navy Workforce Research & Analysis Conference in DC, May 2010
- A Trade Journal or Conference Paper submission in the area of Knowledge Management

DISE

# Relevant Publication

## Facebook Data Set

- "Tastes, Ties, and Time (T3): A new social network dataset using Facebook.com"
  - Authors
    - Kevin Lewisa, Jason Kaufmana, Marco Gonzaleza, Andreas Wimmerb, Nicholas Christakis
  - Affiliations
    - *Department of Sociology, Harvard University, United States Department of Sociology, University of California, Los Angeles, United States*
- Note: T3 dataset ensures privacy of students' profiles

DISE

# Published Research Data

- **1,700 Facebook profiles**
  - **Data set includes**
    - User profiles
    - Interests
    - Communities
    - Forums, group affiliations (schools, colleges and universities, companies, etc.)
    - Friends list, fandom (fan of)
    - Endorsements (supporter of)
- **Navy Enlisted Rating descriptions**
  - *Hard Cards*
  - Officer Community Descriptions

DISE

# Potential Research Benefits

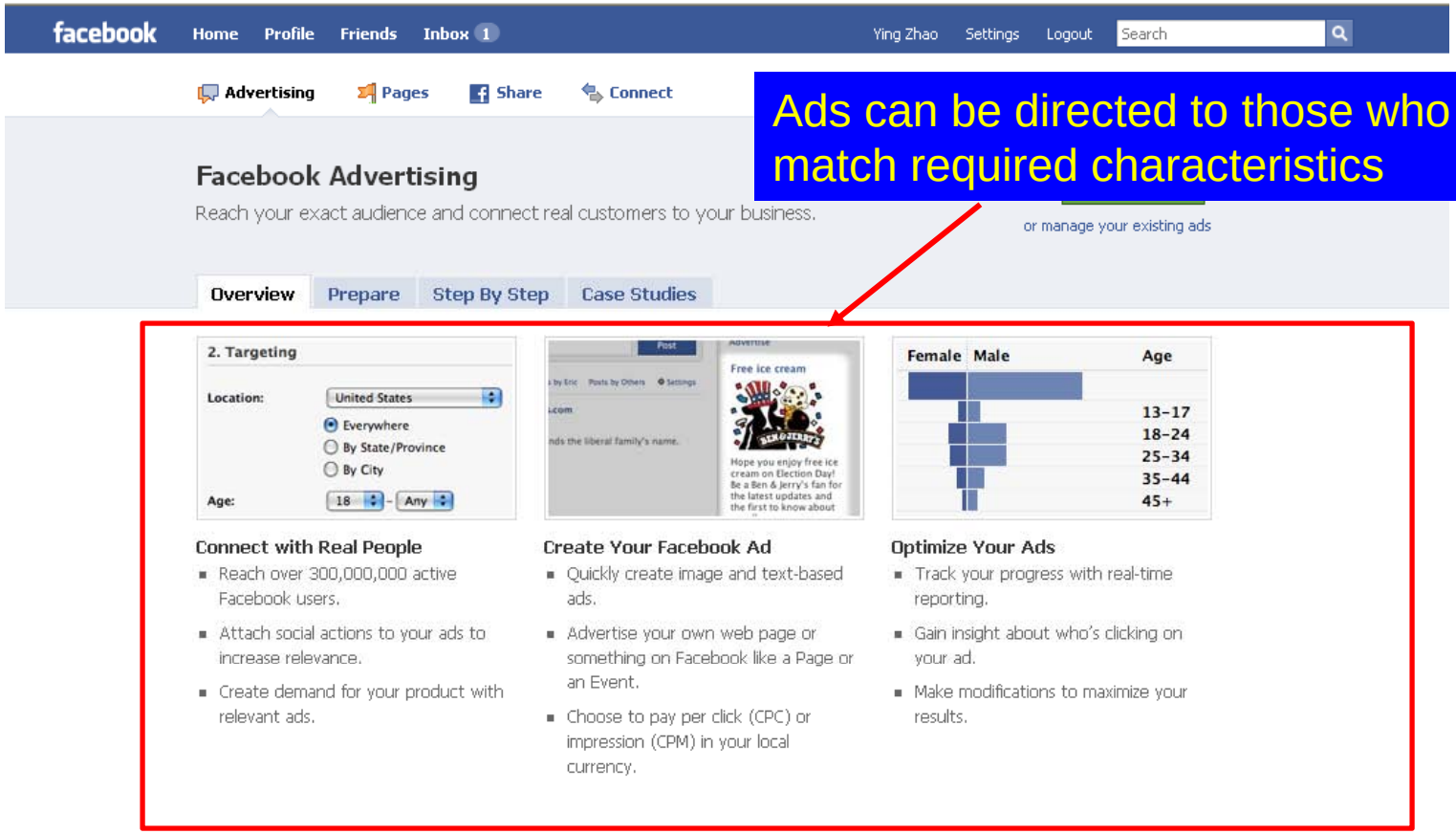
- Improved strategic decisions
- Lexical links can be used to improve
  - Recruiting campaigns
  - Advertisements
  - Regionalization understanding
  - Marketing event strategies
- Real-time marketing
  - Facebook page that takes the feed and disseminates recruiting marketing messages

DISE

# Potential Impact

DISE

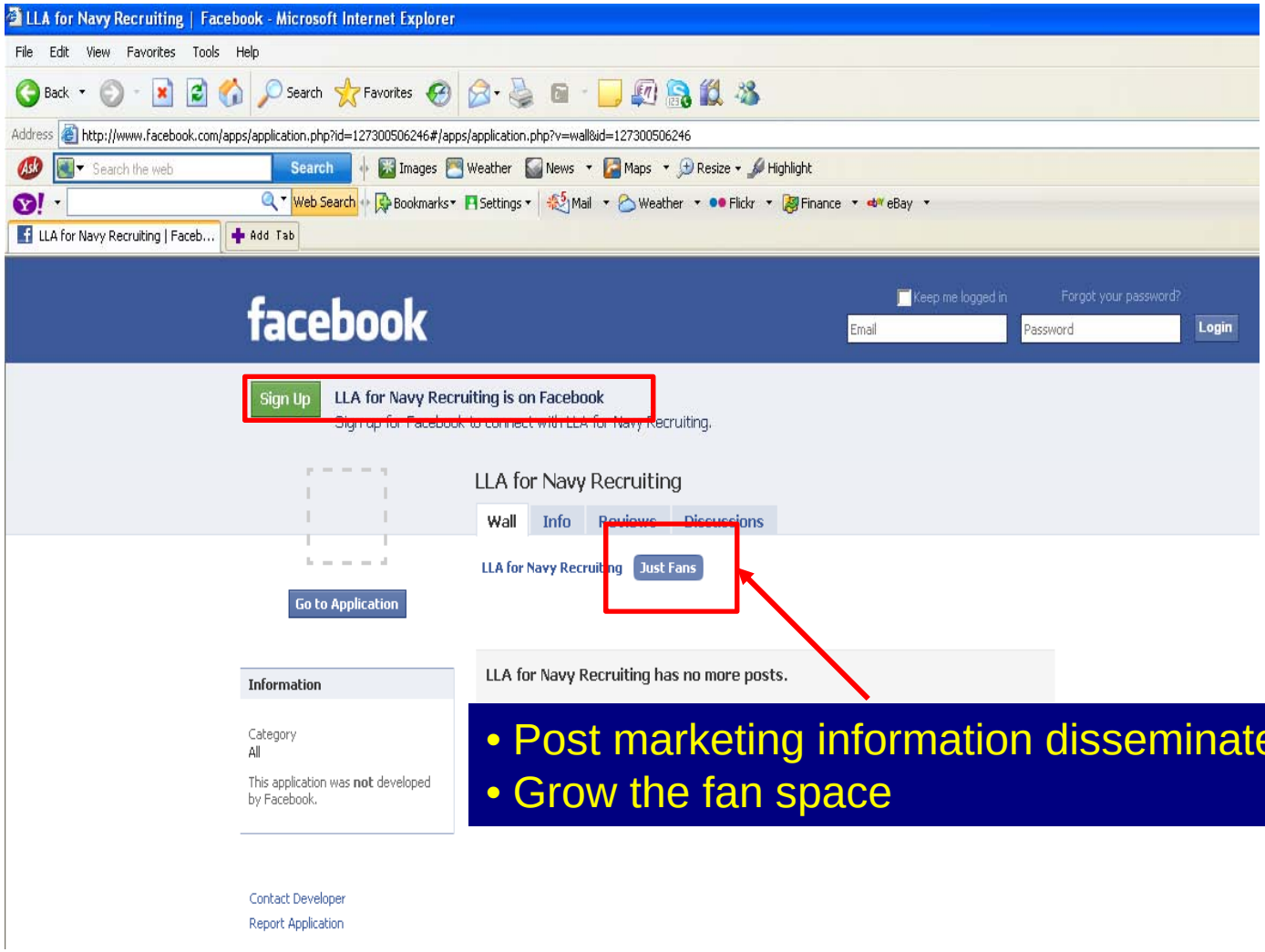
# Direct Advertising Using LLA Results




Ads can be directed to those who match required characteristics

# Create a dedicated application

DISE




- Post marketing information disseminated to fans
- Grow the fan space

DISE

# Ads Placed via Facebook to Matched Profiles & Response Results Tracked

DISE

# Facebook Profile Data Example

**Connect**

**Interests (36):**
adventure, archipelagoes, cavalry, cold war, dragons, fantasy, fiction, geography, heraldry, history, immortality, longevity, love story, magic, military history, mindpowers, mounted combat, nanotechnology, nanowrimo, national novel writing month, naval history, novel writing, parapsy psionics, psychic powers, regeneration, sailing ships, science fiction, storytelling, time travel, tropical islands, warfare, wizardry

**External Services:**
- tanelos@livejournal.com
- rizanabsith
- 191317559
- rizanabsith
- rizanabsith@hotmail.com
- rizanabsith@gotrillian.com
- banazir
- banazir

**Schools:**
Hillsborough Elementary School - Hillsborough, NJ (1983 - 1984)
Scott Lake Elementary School - Lakeland, FL (1984 - 1985)
Severna Park Middle School - Severna Park, MD (1985 - 1987)
Severn School - Severna Park, MD (1987 - 1989)
Johns Hopkins University - Baltimore, MD (1989 - 1993)
University of Illinois at Urbana-Champaign - Urbana, IL (1993 - 1998)

**Friends** [View Entries]

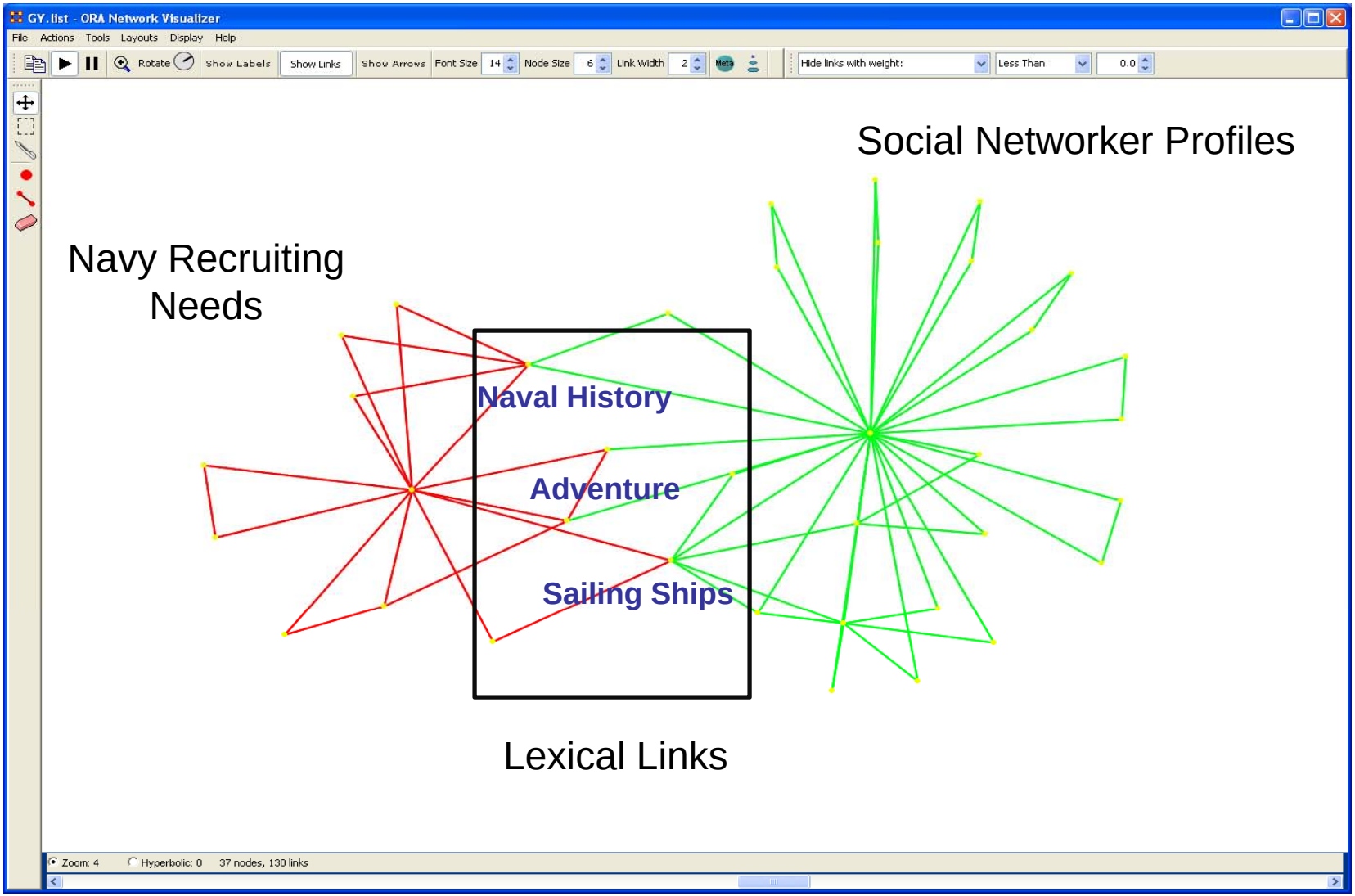
**Friends (66):**
amishlurker, andrewwyldgonzo, anim8t0rboi, banazir, bdm7935, burkhardt, carida_46, cavlec, cheyinka, cloverzj, cretaceousrick, darana, deire, dragnflye, dsthenes, elvenwanderer, figgylicious, gngr, gondhir, gothic_spire, illusio, istari_ala, jadziadax, kaladhwen, klawz, koinonia, ldymlissa masaga, masteralida, megruder, miajesbar, mirabehn, miyeko, mom_counsel, mrowe, myng_rabbyt, narvi, neoookami, nobuddy69, pathofthejus queenmidalah, ringlass, rivendellrose, sahtyinepu, smegalicious, sperose, sui_degeneris, tabula_xrasa, taiji_jian, tamf, teremala, the_wanlorn, t twinofhugin, yodge, zengeneral, zurich31

**2. Data**

| | | |
|---|---|---|
| w1dataset.xls<br>MS Excel - 314 KB - 24 downloads | Restricted Access | wave 1 general attributes |
| w1fbk.net<br>Plain Text - 1 MB - 21 downloads | Restricted Access | wave 1 facebook friends |
| w1hsgrp.net<br>Plain Text - 218 KB - 21 downloads | Restricted Access | wave 1 housing groupmates |
| w1pic.net<br>Plain Text - 357 KB - 21 downloads | Restricted Access | wave 1 picture friends |
| w1room.net<br>Plain Text - 170 KB - 21 downloads | Restricted Access | wave 1 roommates |
| w1tastes.xls<br>MS Excel - 1 MB - 21 downloads | Restricted Access | wave 1 cultural tastes |

DISE

# Notional Lexical Link Map
# Matching Profiles to Navy Ratings



# LLA Recruiting Validation Methods

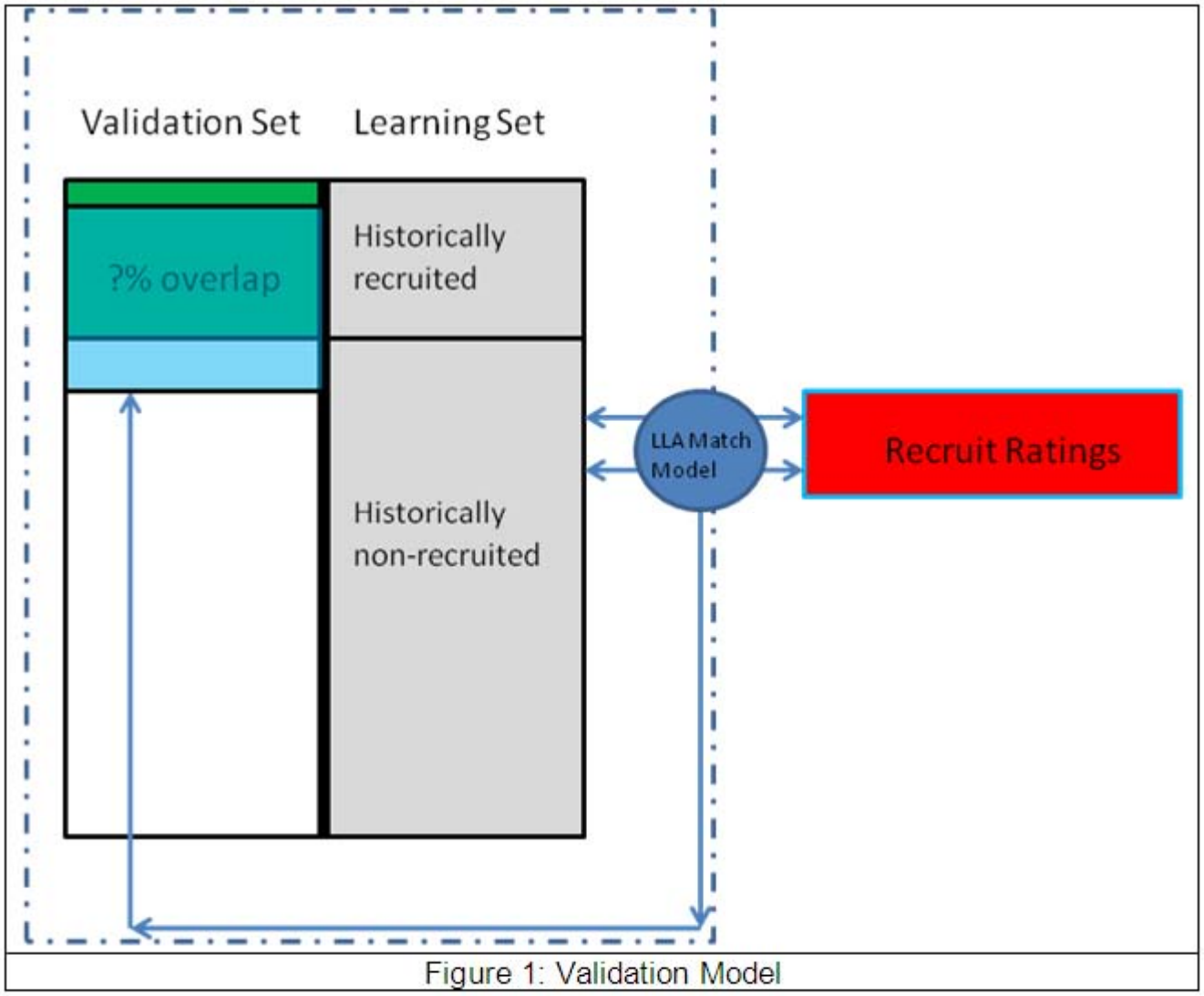



Figure 1: Validation Model

DISE

# LLA Recruiting Validation Method

- **Step 1**: Consider historical sample profiles of recruited personnel, mixed with general populations, for example, samples of youth population in US. The profiles should be similar to the self-reported education and personality characteristics from Facebook.
- **Step 2**: Divide the samples into a validation and a learning set. Set aside the validation set. Use the learning set to match the recruit ratings with the proposed LLA and generate a LLA Match Model.
- **Step 3**: Apply the LLA Match Model to the validation set to recommend the top candidates. Compute the percentage of overlap between the recommended recruits with those who were successfully selected historically (by experts). (Please see the process illustration in Figure 1.) If the overlap percentage or correlation is high, we will consider the matching method as valid.
  - This method can then be used dynamically in automation and thus reduce required manpower.

DISE

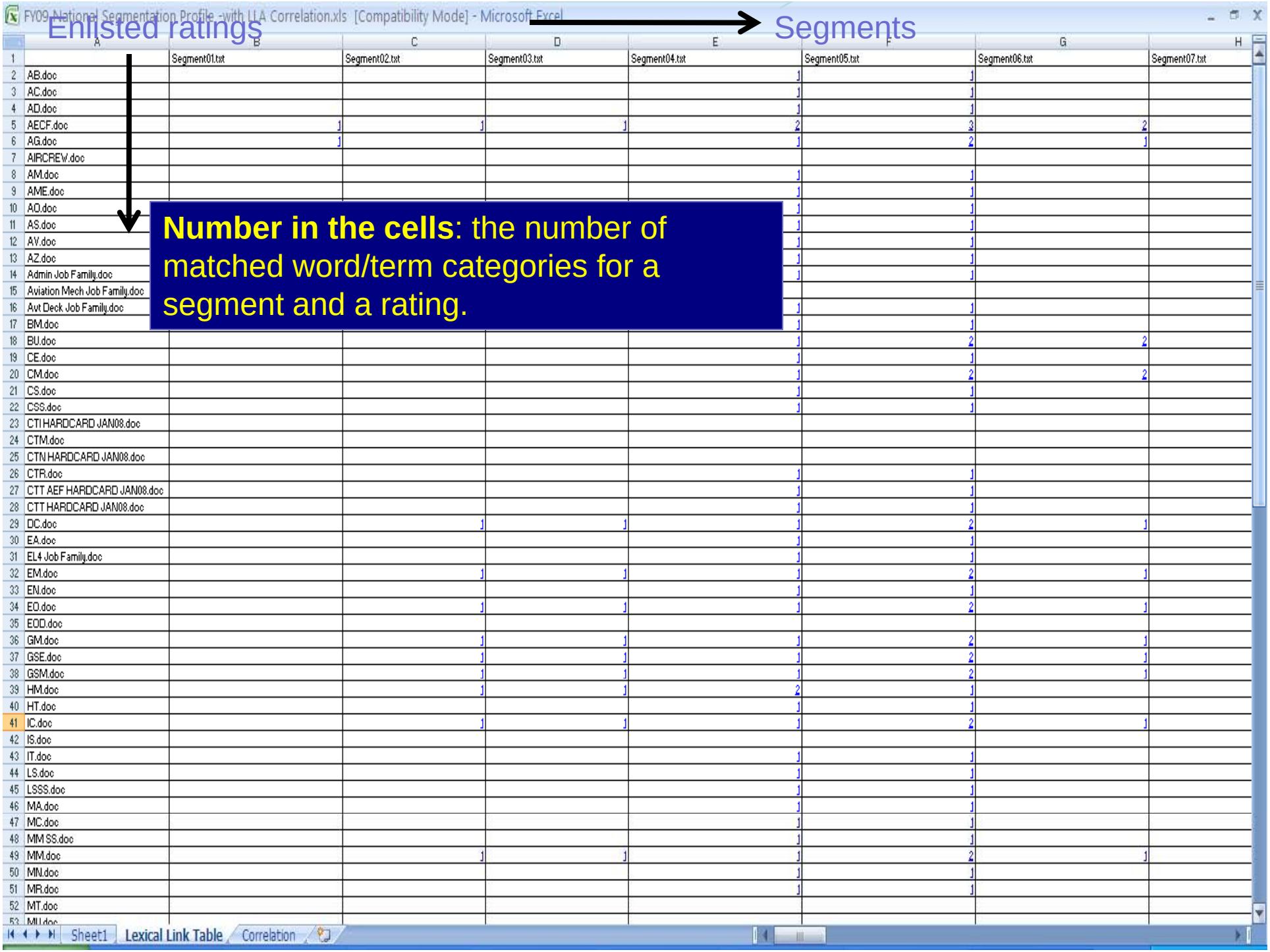
# Rating Descriptions vs. Population Segment

- Data Used
  - Enlisted ratings (AB.doc, AC.doc, AD.doc….)
    - Descriptions
  - JAMRS_2008_SegmentationGuide.pdf
    - 66 segments for 16-20 years of age
- For Validation
  - FY09 National Segmentation Profile.xls
    - Accession rates for all services, for all segments were used for correlation

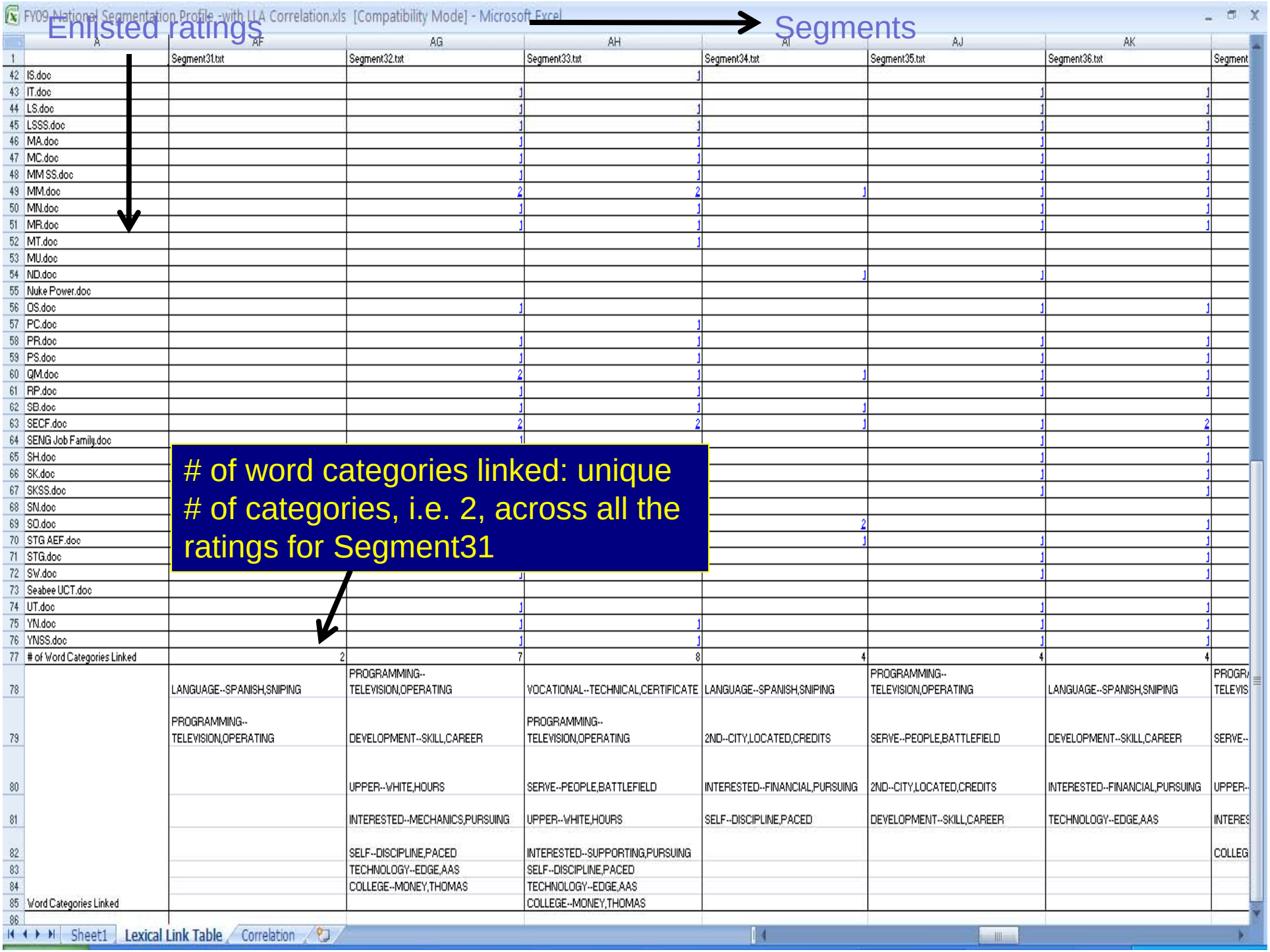
DISE

# Result: Lexical Link Table

- **Rows**: Segment description ids
- **Columns**: Enlisted rating ids
- **Number in the cells**: the number of matched word/term categories for a segment and a rating.



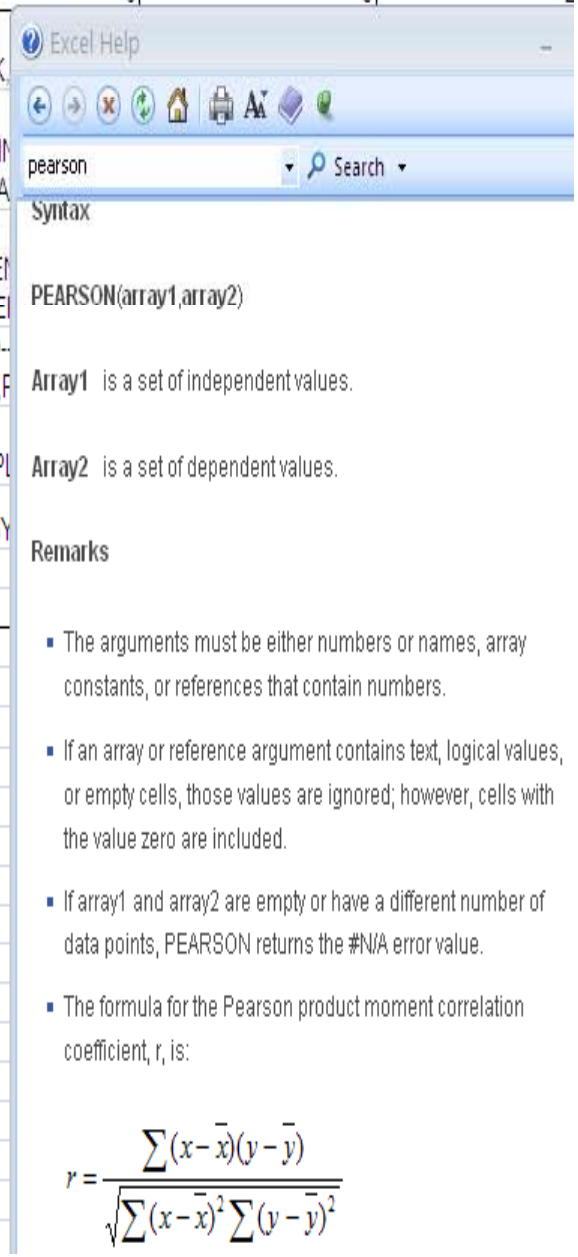

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | | Segment01.txt | Segment02.txt | Segment03.txt | Segment04.txt | Segment05.txt | Segment06.txt | Segment07.txt |
| 2 | AB.doc | | | | 1 | 1 | | |
| 3 | AC.doc | | | | 1 | 1 | | |
| 4 | AD.doc | | | | 1 | 1 | | |
| 5 | AECF.doc | 1 | 1 | 1 | 2 | 3 | 2 | |
| 6 | AG.doc | 1 | | | 1 | 2 | 1 | |
| 7 | AIRCREW.doc | | | | | | | |
| 8 | AM.doc | | | | 1 | 1 | | |
| 9 | AME.doc | | | | 1 | 1 | | |
| 10 | AO.doc | | | | 1 | 1 | | |
| 11 | AS.doc | | | | 1 | 1 | | |
| 12 | AV.doc | | | | 1 | 1 | | |
| 13 | AZ.doc | | | | 1 | 1 | | |
| 14 | Admin Job Family.doc | | | | 1 | 1 | | |
| 15 | Aviation Mech Job Family.doc | | | | | | | |
| 16 | Avt Deck Job Family.doc | | | | 1 | 1 | | |
| 17 | BM.doc | | | | 1 | 1 | | |
| 18 | BU.doc | | | | 1 | 2 | 2 | |
| 19 | CE.doc | | | | 1 | 1 | | |
| 20 | CM.doc | | | | 1 | 2 | 2 | |
| 21 | CS.doc | | | | 1 | 1 | | |
| 22 | CSS.doc | | | | 1 | 1 | | |
| 23 | CTI HARDCARD JAN08.doc | | | | | | | |
| 24 | CTM.doc | | | | | | | |
| 25 | CTN HARDCARD JAN08.doc | | | | | | | |
| 26 | CTR.doc | | | | 1 | 1 | | |
| 27 | CTT AEF HARDCARD JAN08.doc | | | | 1 | 1 | | |
| 28 | CTT HARDCARD JAN08.doc | | | | 1 | 1 | | |
| 29 | DC.doc | | 1 | 1 | 1 | 2 | 1 | |
| 30 | EA.doc | | | | 1 | 1 | | |
| 31 | EL4 Job Family.doc | | | | 1 | 1 | | |
| 32 | EM.doc | | 1 | 1 | 1 | 2 | 1 | |
| 33 | EN.doc | | | | 1 | 1 | | |
| 34 | EO.doc | | 1 | 1 | 1 | 2 | 1 | |
| 35 | EOD.doc | | | | | | | |
| 36 | GM.doc | | 1 | 1 | 1 | 2 | 1 | |
| 37 | GSE.doc | | 1 | 1 | 1 | 2 | 1 | |
| 38 | GSM.doc | | 1 | 1 | 1 | 2 | 1 | |
| 39 | HM.doc | | 1 | 1 | 2 | 1 | | |
| 40 | HT.doc | | | | 1 | 1 | | |
| 41 | IC.doc | | 1 | 1 | 1 | 2 | 1 | |
| 42 | IS.doc | | | | | | | |
| 43 | IT.doc | | | | 1 | 1 | | |
| 44 | LS.doc | | | | 1 | 1 | | |
| 45 | LSSS.doc | | | | 1 | 1 | | |
| 46 | MA.doc | | | | 1 | 1 | | |
| 47 | MC.doc | | | | 1 | 1 | | |
| 48 | MM SS.doc | | | | 1 | 1 | | |
| 49 | MM.doc | | 1 | 1 | 1 | 2 | 1 | |
| 50 | MN.doc | | | | 1 | 1 | | |
| 51 | MR.doc | | | | 1 | 1 | | |
| 52 | MT.doc | | | | | | | |
| 53 | MU.doc | | | | | | | |




| | A | AF | AG | AH | AI | AJ | AK | |
|---|---|---|---|---|---|---|---|---|
| 1 | | Segment31.txt | Segment32.txt | Segment33.txt | Segment34.txt | Segment35.txt | Segment36.txt | Segment |
| 42 | IS.doc | | | 1 | | | | |
| 43 | IT.doc | | 1 | | | 1 | 1 | |
| 44 | LS.doc | | 1 | 1 | | 1 | 1 | |
| 45 | LSSS.doc | | 1 | 1 | | 1 | 1 | |
| 46 | MA.doc | | 1 | 1 | | 1 | 1 | |
| 47 | MC.doc | | 1 | 1 | | 1 | 1 | |
| 48 | MM SS.doc | | 1 | 1 | | 1 | 1 | |
| 49 | MM.doc | | 2 | 2 | 1 | 1 | 1 | |
| 50 | MN.doc | | 1 | 1 | | 1 | 1 | |
| 51 | MR.doc | | 1 | 1 | | 1 | 1 | |
| 52 | MT.doc | | | 1 | | | | |
| 53 | MU.doc | | | | | | | |
| 54 | ND.doc | | | | 1 | 1 | | |
| 55 | Nuke Power.doc | | | | | | | |
| 56 | OS.doc | | 1 | | | 1 | 1 | |
| 57 | PC.doc | | | 1 | | | | |
| 58 | PR.doc | | 1 | 1 | | 1 | 1 | |
| 59 | PS.doc | | 1 | 1 | | 1 | 1 | |
| 60 | QM.doc | | 2 | 1 | 1 | 1 | 1 | |
| 61 | RP.doc | | 1 | 1 | | 1 | 1 | |
| 62 | SB.doc | | 1 | 1 | 1 | | | |
| 63 | SECF.doc | | 2 | 2 | 1 | 1 | 2 | |
| 64 | SENG Job Family.doc | | 1 | | | 1 | 1 | |
| 65 | SH.doc | | | | | 1 | 1 | |
| 66 | SK.doc | | | | | 1 | 1 | |
| 67 | SKSS.doc | | | | | 1 | 1 | |
| 68 | SN.doc | | | | | | | |
| 69 | SO.doc | | | | 2 | | 1 | |
| 70 | STG AEF.doc | | | | 1 | 1 | 1 | |
| 71 | STG.doc | | | | | 1 | 1 | |
| 72 | SW.doc | | 1 | | | 1 | 1 | |
| 73 | Seabee UCT.doc | | | | | | | |
| 74 | UT.doc | | 1 | | | 1 | 1 | |
| 75 | YN.doc | | 1 | 1 | | 1 | 1 | |
| 76 | YNSS.doc | | 1 | 1 | | 1 | 1 | |
| 77 | # of Word Categories Linked | 2 | 7 | 8 | 4 | 4 | 4 | |
| 78 | | LANGUAGE--SPANISH,SNIPING | PROGRAMMING--TELEVISION,OPERATING | VOCATIONAL--TECHNICAL,CERTIFICATE | LANGUAGE--SPANISH,SNIPING | PROGRAMMING--TELEVISION,OPERATING | LANGUAGE--SPANISH,SNIPING | PROGR/TELEVIS |
| 79 | | PROGRAMMING--TELEVISION,OPERATING | DEVELOPMENT--SKILL,CAREER | PROGRAMMING--TELEVISION,OPERATING | 2ND--CITY,LOCATED,CREDITS | SERVE--PEOPLE,BATTLEFIELD | DEVELOPMENT--SKILL,CAREER | SERVE-- |
| 80 | | | UPPER--WHITE,HOURS | SERVE--PEOPLE,BATTLEFIELD | INTERESTED--FINANCIAL,PURSUING | 2ND--CITY,LOCATED,CREDITS | INTERESTED--FINANCIAL,PURSUING | UPPER-- |
| 81 | | | INTERESTED--MECHANICS,PURSUING | UPPER--WHITE,HOURS | SELF--DISCIPLINE,PACED | DEVELOPMENT--SKILL,CAREER | TECHNOLOGY--EDGE,AAS | INTERES |
| 82 | | | SELF--DISCIPLINE,PACED | INTERESTED--SUPPORTING,PURSUING | | | | COLLEG |
| 83 | | | TECHNOLOGY--EDGE,AAS | SELF--DISCIPLINE,PACED | | | | |
| 84 | | | COLLEGE--MONEY,THOMAS | TECHNOLOGY--EDGE,AAS | | | | |
| 85 | Word Categories Linked | | | COLLEGE--MONEY,THOMAS | | | | |
| 86 | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | ALL SVCS Accessions Rate | 0.003824919 | 0.00706807 | 0.006161745 | 0.005987297 | 0.008384598 | 0.009651198 | 0.00588802 |
| 3 | # of Word Categories Linked | 2 | 2 | 2 | 3 | 6 | 5 | 2 |
| 4 | | PROGRAMMING--READ,OPERATING | SERVE--PEOPLE,BATTLEFIELD | SERVE--PEOPLE,BATTLEFIELD | SERVE--PEOPLE,BATTLEFIELD | PART--WORK, | | |
| 5 | | INTERESTED--MECHANICS,PURSUING | SELF--DISCIPLINE,PACED | SELF--DISCIPLINE,PACED | DEVELOPMENT--SKILL,CAREER | PROGRAMMI READ,OPERA | | |
| 6 | | | | | INTERESTED--MECHANICS,PURSUING | DEVELOPME SKILL,CAREE | | |
| 7 | | | | | | INTERESTED-- MECHANICS,F | | |
| 8 | | | | | | SELF--DISCIPL | | |
| 9 | | | | | | TECHNOLOGY | | |
| 10 | | | | | | | | |
| 11 | Word Categories Linked | | | | | | | |
| 12 | | Correlation | | | | | | |
| 13 | | 0.435522936 | | | | | | |

- Pearson correlation between **"# of Word Categories Linked"** and **the accession rate for "ALL SVCS Accessions"** from FY09 National Segmentation Profile.xls.
- The correlation of Row 2 vs. 3 is 0.43 (on a scale of 0 to 1)
- Provides Initial validation for this methodology



DISE

# Conclusions

- Serves as initial validation LLA method
- We expect that more detailed, individual profiles (e.g. facebook) will increase this correlation
- Improve the accession rate in real-life
- Word/term matches can be used to fit different segment youth to different ratings

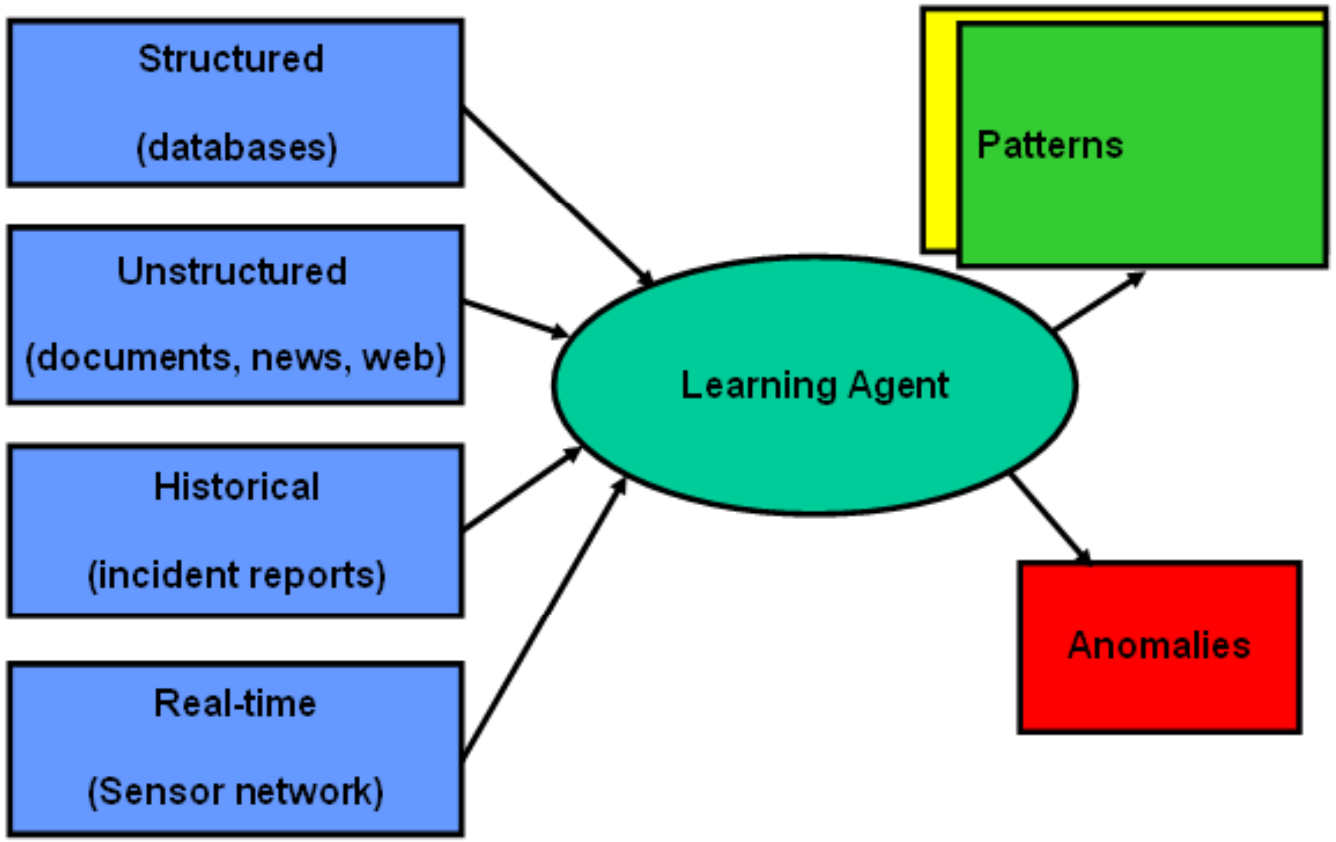
DISE

# References

DISE

# Collaborative Learning Agents (CLA)

(http://www.quantumii.com)



DISE

# Internet References

- http://blog.jobzcafe.com/2008/09/08/facebook-for-recruiting-whats-working/
- http://www.facebook.com/pages/US-Navy-SEAL-SWCC-EOD-Diver-Rescue-Swimmer-recruiting/166379650511?v=feed&story_fbid=247595570511&ref=mf
- http://www.facebook.com/pages/Chambersburg-PA/Navy-Recruiting-Station-Chambersburg-PA/116537271249